Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 521 – 530

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

# Improving the Connectivity of Community Detection-based Hierarchical Routing Protocols in Large-Scale WSNs

Matheus A. de Paulo[a], Mariá C. V. Nascimento[a], Valério Rosset[a,*]

[a]*Federal University of São Paulo, Avenida Cesare Mansueto Giulio Lattes, 1201, Parque Tecnológico, 12247014, São José dos Campos-SP-Brazil*

**Abstract**

The recent growth in the use of wireless sensor networks (WSNs) in many applications leads to the raise of a core infrastructure for communication and data gathering in Cyber-Physical Systems (CPS). The communication strategy in most of the WSNs relies on hierarchical clustering routing protocols due to their ad hoc nature. In the bulk of the existing approaches some special nodes, named Cluster-Heads (CHs), have the task of assembling clusters and intermediate the communication between the cluster members and a central entity in the network, the Sink. Therefore, the overall efficiency of such protocols is highly dependent on the even distribution of CHs in the network. Recently, a community detection-based approach, named RLP, have shown interesting results with respect to the CH distribution and availability that potentially increases the overall WSN efficiency. Despite the better results of RLP regarding the literature, the adopted CH election algorithm may lead to a CH shortage throughout the network operation. In line with that, in this paper, we introduce an improved version of RLP, named HRLP. Our proposal includes a hybrid CH election algorithm which relies on a computationally cheap and distributed probabilistic-based CH recovery procedure to improve the network connectivity. Additionally, we provide a performance analysis of HRLP and its comparison to other protocols by considering a large-scale WSN scenario. The results evince the improvements achieved by the proposed strategy by means of the network connectivity and lifetime metrics.


Peer-review under responsibility of KES International

*Keywords:*
Community detection; routing; wireless sensor networks.

## 1. Introduction

Wireless Sensor Networks (WSNs) are ad hoc networks that may be composed by hundreds to thousands resource-constrained sensor nodes[1]. In such networks, the task of each sensor node is to deliver previously acquired data from the environment to a resourceful node, the Sink. This procedure is attainable by built-in wireless communication capabilities, such as those provided by low-powered RF devices.

As stated by Santi[2], collaboration among sensor nodes is essential to minimize and evenly distribute the impact that communication costs have in the energy availability. Hierarchical routing protocols have been extensively adapted and

* Corresponding author. Tel.: +55-12-3924-9500 ; fax: +55-12-3924-9500.
*E-mail address:* vrosset@unifesp.br


Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.229

developed over the last decade[3]. This trend could be justified by their ability to provide simple, yet efficient clustering routing protocols for low to medium scale WSNs. In hierarchical clustering routing protocols, some elected nodes are responsible for assembling clusters by gathering other nodes within their communication range. These cluster-head nodes (CHs) intermediate the communication between the cluster members and the Sink while performing tasks such as cluster coordination and data aggregation. Due to the high communication costs of long-range transmissions, the CHs demand more energy than other nodes. Heinzelman et al.[4] state that the task of CH must be evenly distributed among sensor nodes to ensure energy-efficiency. Accordingly, the great majority of hierarchical clustering routing protocols employ, in some extent, a periodical CH task rotation among the available sensor nodes. The network operation is thus split in a sequence of time-intervals (named rounds), where each of them is preceded by a CH election.

To reduce the computational complexity of the cluster formation, some protocols only allow CHs to gather nodes within their one-hop neighborhood. This can lead to a disconnection of some nodes of the WSN. Consequently, situations in which there is no path between the sensor nodes and the Sink may happen and disrupt the communication in the WSN. This situation happens specially when the sensor nodes are not able to increase their transmission range. Therefore, when considering applications that expect strict response times from sensor nodes, such as tracking and intruder-detection, the overall efficiency of hierarchical clustering routing protocols is tightly bounded to their ability of keeping the even distribution of CHs in every round. Hierarchical clustering protocols were classified by Mamalis et al.[5] according to their CH election algorithms. Probabilistic-based protocols employ probability distributions to produce fast and low complexity CH election algorithms. Meanwhile, non-probabilistic protocols rely on graph-based or distributed heuristic algorithms to pursue the optimal CH distribution. The authors of[6] applied a community detection-based routing protocol to balance the CH distribution between the network communities. They have shown that this approach can lead to a more even CH distribution than the strategy proposed in[4]. However, a fading in the availability of CHs over the elapse of rounds was observed. This behavior results from the CH election procedure adopted in[6], which relies on an indication of the next CHs by the actual CHs into communities.

In line with the low complexity of probabilistic-based CH election algorithms and the observed deficiency of community detection-based CH election described in[6], this paper introduces a hybrid CH election algorithm for community detection-based WSN routing protocols. The proposed approach extends the original CH election algorithm presented in[6] by including a distributed probabilistic-based CH recovery procedure which improves the connectivity between nodes and CHs inside communities.

The remaining sections are organized as follows. Section 3 presents a survey of some noteworthy CH election algorithms described in the literature. Section 4 describes the protocol introduced in De Paulo et al.[6], summarizing its main proprieties and adversities while Section 5 describes the CH recovery procedure that improves its CH availability throughout the rounds.

## 2. Notations and Definitions

This section presents the definitions and notations used thorough the paper.

Let $G = (V, E)$ be an undirected graph. The set of vertices (nodes) $V$ is composed by $n$ elements represented by consecutive integers from 1 to $n$. The set of edges, $E$, indicates the pairwise relationship between members of $V$. Therefore, if $(i, j) \in E$, it means that vertices $i$ and $j$ are adjacent in graph $G$. The degree of a node $i$ consists in the number of vertices adjacent to it. Here, the degree of a node $i$ is denoted by $\delta_i$. The average degree of $V$, $\bar{\delta}$, is the ratio between sum of the degrees of every vertex from $V$ and $n$.

The neighborhood of a vertex $i$, i.e., the set of neighbors of $i$, is $N_i = \{j \in V : (i, j) \in E\}$. A clique is a set of pairwise adjacent vertices. A $k$-partition of its vertex set is some finite collection $C = \{C_1, C_2, \ldots, C_k\}$ where each $C_j \in C$ is a set or community of highly related vertices. Equations (1) and (2) describe the main properties of communities in $G$.

$$\bigcup_{i=1}^{k} C_i = V \tag{1}$$

$$C_i \cap C_j = \emptyset, \quad \forall i \neq j. \tag{2}$$

## 3. Related Work

In this section we present the CH election algorithms most related to our proposal. The classification of these algorithms into probabilistic-based and non-probabilistic-based is according to Mamalis et al.[5].

### 3.1. Probabilistic-based CH election algorithms

Probabilistic-based CH election approach relies on low complexity probability distributions. In[4], the authors present the Low-Energy Adaptive Clustering Hierarchical (LEACH) protocol. Using a simple distribution probability, their CH election algorithm ensures that an expected amount of CHs elected in every round can be previously defined by a global parameter. However, the expected distribution of CHs at a given round may not guarantee any level of network connectivity. This occurs because, although CHs are randomly elected among the sensor nodes, the election approach does not provide guarantees about the even distribution of such elected CHs over the network.

Equation (3) presents the core mechanism of the CH election employed by LEACH. It calculates the probability $P$ of a node $i$ to be elected as CH in round $r$. In this equation, $p$ is the expected proportion of CHs in any given round and $C_r$ is the set of nodes that have not been a CH in the last $\left\lceil \frac{1}{p} \right\rceil$ rounds. Therefore, every node in the network calculates $P$ and becomes a CH for a given round if and only if the value of $P$ is greater than a value chosen by chance in $[0, 1]$.

$$P(r) = \begin{cases} \frac{p}{1-p\left(r \mod \frac{1}{p}\right)}, & \text{if } i \in C_r, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The load balance and energy-efficiency strategy of LEACH relies only on assigning the task of CH to sensor nodes that were not elected on the former $\left\lceil \frac{1}{p} \right\rceil$ rounds. Therefore, Handy et al.[7] adapted the CH distribution probability designed in[4] to take into account the energy availability of each sensor node. Their expected result is that nodes with low residual energy would have a small probability of been selected as CH. Equation (4) presents the probability $P'$of node $i$ becomes a CH in round $r$ as defined in[7].

$$P'(r) = \frac{p}{1 - p\left(r \mod \frac{1}{p}\right)} \left[e_i + K_{ir}\left(1 - e_i\right)\right]. \tag{4}$$

In Equation (4), $e_i \in [0, 1]$ is the residual energy of node $i$ and $K_{ir}$ is a factor proportional to the number of consecutive rounds that node $i$ was not a CH, Handy et al.[7] claim that the convex combination introduced in Equation 4 enables a balance between energy conservation and the CH availability. However, likewise in[4], the strategy adopted in[7] does not guarantee any level of network connectivity. Bearing that in mind, Liu et al.[8] introduced an adaptation of the election strategy of[4] considering the ratio between the degree of the node candidate to be a CH and the average degree of nodes of the network, the relative degree.

Additionally, the strategy also takes into account the residual energy of each sensor node. Equation (5) presents the probability $P''$ of node $i$ be elected as CH in round $r$ calculated by sensor nodes in the protocol proposed in[8], named LEACH-D.

$$P''(r) = \frac{p}{1 - p\left(r \mod \frac{1}{p}\right)} e_i \frac{\delta_i}{\bar{\delta}}. \tag{5}$$

The use of the relative degree in Equation (5) enables a CH distribution that provides a good connectivity at the beginning of the protocol execution. The reason behind that is that nodes with the highest relative degrees can clearly gather more nodes into their clusters. Nevertheless, the CH availability of LEACH-D may decrease over the passing of rounds, as the energy consumption tends to be more intense in those sensor nodes with the highest relative degrees. Therefore, leading them to exhaust their energy supply faster than other sensor nodes. Besides that, the maintenance of relative degree information by sensor nodes is costly to the network since they need to periodically to exchange control messages.

### 3.2. Non-probabilistic-based CH election algorithms

Although probabilistic-based CH election algorithms have low computational cost, they produce random CH distributions among the network. Differently, the non-probabilistic-based CH election algorithms are mostly heuristics and that aim at better distributing the CHs among the network according to a given criterion. Chang and Perrig[9] proposed an adaptive strategy that iteratively *spawns* and *migrates* clusters in the designed of Algorithm for Cluster Establishment (ACE). Each CH election procedure in ACE is split in $c$ asynchronous iterations. Nodes declare themselves as CH in iteration $i$ if $l \geq f_{min}(i)$, where $l$ is the number of neighboring nodes that are not members of any other cluster (named, loyal followers) and $f_{min}(i)$ is the spawning threshold at iteration $i$, given by the Equation (6), where $k_1$ and $k_2$ are predefined constants.

$$f_{min}(i) = \left(e^{-k_1 \frac{i}{c}} - k_2\right) \bar{\delta}. \tag{6}$$

In ACE, initially the CHs recruit all neighbors to their clusters, even those that already belong to other clusters. Later, a cluster migration is performed by every CH to reduce the number of members overlapping with other clusters. In the migration process each CH seeks for a member of its cluster with the greatest value of $l$. If such member does exist it is promoted as a new CH, otherwise the CH remains the same. Therefore, the cluster members that are adjacent to the promoted CH are kept in the cluster, whereas those which are not can be part of other cluster.

Likewise in ACE, Wen and Sethares[10] introduced the Clustering Algorithm via Waiting Timer (CAWT) aiming at a fast and distributed CH election algorithm with the optimal connectivity. CAWT relies on the use of adaptive waiting times before a node decides to become a CH or join a nearby cluster. At the protocol beginning every sensor node $i$ initializes a random waiting timer variable ($WT_i$). In addition, each sensor node broadcasts a *hello* message after $\lambda\ WT_i\ U(0, 1)$ units of time, where $\lambda \in (0, 0.5)$ and $U$ is a uniform distribution. In the mean time, whenever a sensor node $i$ receives a *hello* message from any of its neighbors, the value of $WT_i$ is reduced by a factor of $\beta \in (0, 1)$. Therefore, if the $WT_i$ expires and none of $i$ neighbors belong to any cluster then $i$ declares itself as a CH by sending a CH announcement message to all its neighbors. In this case, if any sensor node receives a CH announcement, it becomes a member of the announced cluster and no longer attempts to elect itself as CH.

Brust et al.[11] introduced a CH election algorithm based on a strong local connectivity parameter. The clustering coefficient of node $i$ ($CC_i$), given by Equation (7), measures how far its neighborhood is from been a clique[12]. Each node only requires information from its 2-hop neighborhood to evaluate its clustering coefficient, allowing for a attainable local metric in WSNs.

$$CC_i = \frac{2\,|\{(u, v) \in E : u, v \in N_i\}|}{|N_i|\,(|N_i| - 1)}. \tag{7}$$

In order to efficiently elect CHs in a given network, the strategy proposed in[11] first classifies each node according to the following. Nodes with less than three neighbors or with a clustering coefficient lower than a threshold are classified as *weak nodes*. Otherwise, they are classified as *strong nodes*. Afterward, nodes represented by cut vertices in $G$ are classified as *bridge nodes* since they represent single points of failures by causing a network disconnection. Furthermore, the strategy also identifies the nodes that are localized on border regions of possible clusters (named *the border nodes*) using a local procedure that checks for two-hop neighborhood intersection at each node. At the end of the election strategy the remaining strong nodes are classified as CHs.

The topology-based CH election algorithms presented so far may guarantee even distribution but not energy-efficiency. In line with that Ngo et al.[13] introduced the Messaging Passing (MEPA) clustering protocol, a distributed preference-based optimization aiming to energy-efficiency and connectivity. According to[13], nodes with the highest residual energy and degree are preferable to play the CH role. Equation (8) presents the normalized preference metric of node $i$ with respect to node $j$.

$$p_i(j) = \frac{e_j}{\sum_{k \in N_i} e_k} \tag{8}$$

In MEPA, the likelihood of a node becoming a CH is proportional to the normalized preferences that it receives from its neighbors, that is from $\{p_i(j) : j \in N_i\}$. In addition, this probability takes into account its self-normalized preference, given by $p_i(i)$. In order to self-elect as CHs, nodes share their preferences with their neighborhood using two types of messages. Request messages $req_i(j)$ sent from node $i$ to neighbor $j \in N_i$ contain the likelihood of $j$ been selected as CH according to $i$. Each request message has a counterpart in a response message $res_i(j)$ sent from node $j$ to node $i$. Equations (9) and (10) evaluate, respectively, the request and response messages from node $i$ to node $j$.

$$req_i(j) = p_i(j) - \max_{\{k \in N_i : k \neq j\}} \{p_i(k) + res_i(k)\}. \tag{9}$$

$$res_i(j) = \begin{cases} \min\{0, req_j(j) \\ + \sum_{\{k \in N_i : k \neq j\}} \max(0, req_j(k))\}, & \text{if } j \in N_i, \\ \sum_{k \in N_i} \max\{0, req_i(k)\}, & \text{if } i = j. \end{cases} \tag{10}$$

After the request and response procedure, nodes are elected as CHs if they maximize a function of normalized preferences and responses within their neighborhood. Equation (11) summarizes the decision factor of the CH election in MEPA.

$$CH_i = \arg\max_{j \in N_i \cup \{i\}} \{res_i(j) + p_i(j)\}. \tag{11}$$

Routing Label Propagation (RLP)[6] is a cluster-based hierarchical protocol that aims to evenly distribute CHs in WSNs. RLP employs a topology-based CH rotation between nodes from the same community, which is detected by a distributed algorithm, named Vertex Label Propagation (VLBP). Although the protocols described in this section have been developed aiming at achieving their optimal criteria, such as connectivity and energy-efficiency, they do not employ any fault tolerance mechanism to recover from CH and communication failures. In the following sections we present a CH recovery procedure designed to improve the performance of the RLP protocol proposed in[6]. For this reason, a general guideline to the VLBP and the CH election procedure of RLP are presented in next section, introducing their main functionality.

## 4. Routing Label Propagation

### 4.1. Vertex Label Propagation Algorithm

In order to detect a vertex partition, De Paulo et al.[6] introduced the Vertex Label Propagation (VLBP) algorithm as a distributed adaptation of the well-established community-detection algorithm Label Propagation (LP)[14]. They considered the following assumptions to represent the WSN as an undirected graph $G$:

- Each sensor node has a unique representation as a vertex of $V$.
- The edge set $E$ represents the local communication capabilities between pairs of devices. That is, $(u, v) \in E$ iff $u, v \in V$ can communicate with each other through the wireless medium.
- The communication radius is unique and shared among sensor nodes. Therefore, each sensor node $i \in V$ has $N_i = \{j \in V : (i, j) \in E\}$ as its neighborhood set.

VLBP is an iterative algorithm based on a joint decision among the neighborhood of nodes to define their label. For this, during the iterations *temporary* or *definitive* labels are assigned to the nodes, depending on the labels of the nodes neighboring them.

The iterations of VLBP are defined by two procedures. Foremost, a label exchanging procedure occurs between neighboring nodes. Each node broadcasts a message carrying its current label and its status (*temporary* or *definitive*). Upon receiving these messages, neighboring nodes store the received labels into separated sets according to their classification. In the end of the label exchanging procedure, nodes that broadcast a *definitive* label halts in the VLBP algorithm. After the label exchanging procedure, the nodes that are still running the VLBP algorithm elect as their new label the most frequent label stored. The *temporary* labels are discarded and nodes classify their new label as *temporary* if the current and previous labels differ, otherwise the new label is set as *definitive*. Each iteration of VLBP requires only one broadcast by node. According to Raghavan et al.[14], the LP algorithm requires a constant number of iterations to converge. Therefore the VLBP also requires a constant number of broadcasts by each node throughout its execution (5 on average as observed in the experiments).

The result of the VLBP algorithm in every node is a *definitive* label representing its community. Figure 1 presents an example of the communities organization after the execution of the VLBP in a WSN composed by 3000 nodes. It is noteworthy that labels only represent the association of nodes with their communities, not guaranteeing that the labels uniquely represent a community in the partition. Besides that, the VLBP execution does not rely on the previous definition of the partition size, resulting in an adaptive community-detection algorithm.

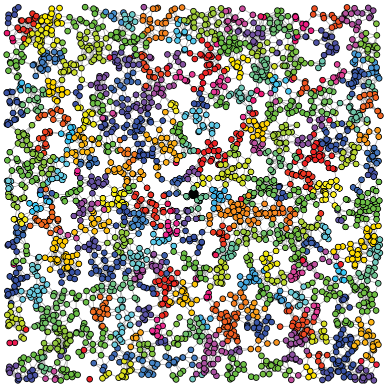


Fig. 1. An example of a 3000 nodes WSN organized in communities generated by the VLBP.

### *4.2. Cluster-Head Election*

In the RLP, the CH task is rotative among the members of the communities. Therefore, in order to assign the role of CH to members in the communities, RLP defines an intra-community CH election algorithm, namely Cluster-Head Establishment (CHE) procedure. At the first round of CHE the set of CHs is composed by the set of Community Border Members (CBMs). A CBM is any node adjacent to at least one node outside its community. In the subsequent rounds, each CH randomly chooses a non-CH node member of its community as the new CH. Every CH then sends an indication message to the respective elected node informing about its new role. The drawback of the CHE procedure is the reduction of CHs caused whenever either a node is elected by more than one CH or the transmission of any indication message fails. Consequently, RLP may experience a shortage of CHs throughout the rounds specially in large-scale WSNs. Therefore, a distributed adaptive CH recovery procedure is proposed in next section to preserve the CH availability during the network operation.

## 5. The proposed CH recovery procedure

As described in Section 3, the probabilistic distributions are suitable for producing computationally cheap CH election algorithms. Therefore, a probabilistic-based CH recovery procedure is proposed in this Section as an extension of the CHE procedure adopted in RLP. The main objective of this extension is to enhance the number of available CHs and therefore, the network connectivity. An exponential distribution describes the elapsed time between occurrences of Poisson-distributed events. One of such events could be interpreted as the CH availability at a given round. In this case, its cumulative exponential distribution could express the probability of a self-election as CH nodes for those nodes that were temporarily disconnected from the Sink.

However, the estimation of Poisson parameters is highly dependent of the current CH distribution. Since in RLP, nodes only operate with local information, such estimation is not trivial. Therefore the exponential-based distribution here introduced was obtained and verified by simulation. If $k \in \mathbb{N}$ is the quantity of consecutive rounds where node $i \in V$ was not a CH or was not connected to any CH, Equation (12) introduces the probability, $\psi$, of a node $i$ to self-elect as a new CH at the end of each CHE procedure of RLP.

$$\psi_i(k, \lambda) = 1 - e^{-\lambda k} \quad (12)$$

where $e^{\lambda} \cong |V|^{\frac{|N_i|}{|V|}}$.

The CH recovery procedure combined with the original CHE procedure of RLP produces a hybrid CH election algorithm, and, therefore an improved version of RLP, here named Hybrid Routing Label Propagation (HRLP).

## 6. Performance Evaluation

A simulation-based evaluation was conducted for evaluating the overall improvement of the proposed HRLP in comparison to the strategies proposed in[4] and[6]. The simulation models were implemented in Castalia/OMNeT++ simulation framework[15].

### 6.1. Evaluation Metrics

For evaluating the described CH election strategies and compare them to each other, as defined in[6], we considered two metrics: the Connectivity and the Lifetime.

**Definition 1** (Connectivity)**.** *The observed connectivity at a given round is the proportion of sensor nodes capable of routing data to the Sink. Therefore, the connectivity is the ratio of the sum of CHs and cluster members to the* $|V|$.

**Definition 2** (*Lifetime*)**.** *The* lifetime *is the amount of rounds that the percentage of clustered nodes (CH and members) remains above a certain connectivity threshold* $\alpha \in [0, 1]$ *for at least* $\gamma \in \mathbb{N}$ *consecutive rounds.*

Therefore one may notice that the better the CH distribution in the network, the greater the network lifetime. The connectivity threshold $\alpha$ measures the desired fraction of the network that must be connected to the Sink. By performing the CH rotation, the effective fraction of connected nodes may vary during the rounds. The recovery threshold $\gamma$ indicates the tolerance, in rounds, of the protocols to the temporary absence of network connectivity.

### 6.2. System Model and Simulation Parameters

For conducting the presented performance analysis we assume a large-scale WSN with the following properties:

- Nodes are homogeneous with respect to hardware capabilities and do not support any level of mobility.
- Nodes admit two distinct levels of transmission power. As in[6] and[4], the lowest level is used for short range communications between nodes and their neighborhood whereas the highest level is used only by CHs to transmit data directly to the Sink.
- All nodes share a wireless communication channel where the medium access control is performed by a Carrier Sense Multiple Access protocol with collision avoidance (CSMA-CA). However, transmissions may fail due to the hidden terminal problem.

Table 1 summarizes the main simulation parameters. The presented simulation results represent the average of values obtained by 40 independent simulation executions of every scenario derived of each parameter arrangement. In every scenario nodes were randomly placed in the environment whilst the Sink was placed in the middle of the field. An example of the nodes placement is illustrated in the Figure 1.

The $p$ parameter of Equation (3) was set to the most frequent CH availability achieved by the HRLP to ensure an equivalent operation of LEACH when compared to HRLP. Therefore, the directed comparison of LEACH and RLP can not be done by observing the results presented in this paper. For this specific comparison, see[6].

### 6.3. Results and Discussion

#### 6.3.1. Connectivity and Lifetime

Figure 2 presents the observed connectivity throughout the rounds of LEACH, RLP and HRLP. It is noteworthy that RLP presents a lower connectivity in comparison to the results achieved by LEACH and HRLP. Despite that, as

Table 1. Simulation Parameters

| Parameter | Values |
|---|---|
| $\|V\|$ (units) | 3000 |
| Environment Area extent ($m^2$) | $868 \times 868$ |
| RF device | CC2420 |
| Lower Transmission Power ($dBm$) | $-5$ |
| Higher Transmission Power ($dBm$) | 0 |
| Simulation Time (rounds) | 1800 |
| $\alpha$ ([0, 1]) | $\{0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ |
| $\gamma$ (rounds) | $\{1, 2, 3, 5, 8, 10, 15, 20, 25, 30\}$ |

expected, LEACH presented a greater variation in the connectivity than HRLP. These results indicate a substantial growth and stability in the connectivity levels achieved by the HRLP in comparison to those obtained in RLP. Additionally, since the proposed CH recovery procedure is computationally cheap it may be adapted to other protocols without compromise their design and operation. Again, it is very important for applications that require strict response times to keep constantly a high connectivity. Bearing that in mind, let us analyze the lifetime results displayed in Figure 3.

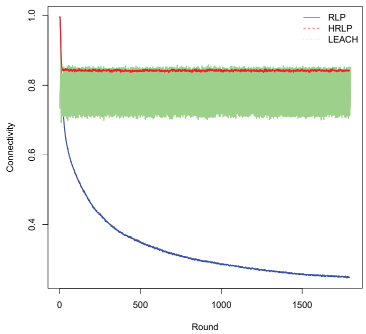


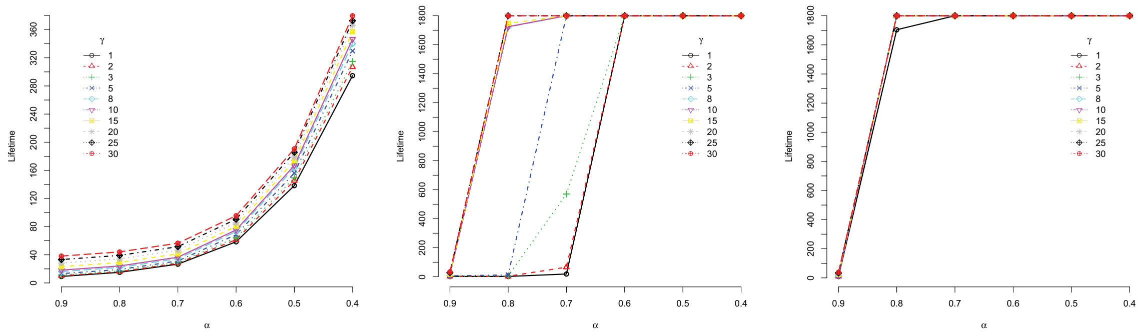

Fig. 2. Connectivity of RLP and HRLP. Each data series represents the connectivity of the CH election algorithms throughout the rounds.

The RLP lifetime results, showed in Figure 3.a, evince its inability of recovering from the reduction in the number of CHs. This reduction is caused by either transmission failures in indication messages or multiple messages indicating the same node as CH. Differently, LEACH achieved the best lifetime results for values of $\gamma$ from 8 to 30 (Figure 3.b). These results support the claim that LEACH may not guarantee constant connectivity. Besides that, the observed small variability in the connectivity of HRLP (Figure 2) is responsible for its best results of *lifetime* even for small values of $\alpha$ as observed in Figure 3.c.

### *6.3.2. CH Distribution and Load Balancing Capabilities*

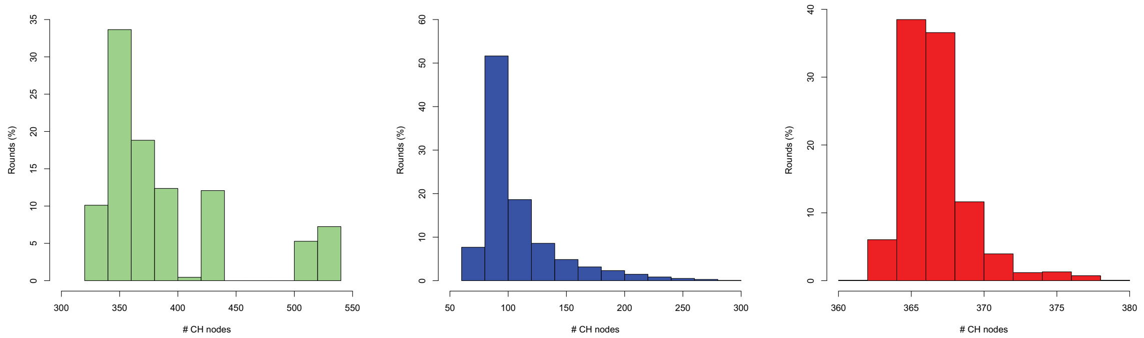
Figure 4 presents histograms related to the CH availability achieved by every evaluated protocol. Each bar of the histogram represents the percentage of rounds with the related range of available CHs. In Figure 4.a, one may observe that although LEACH has presented approximately 386 CHs per round in average, in almost 25% of the rounds the achieved number of CHs significantly deviates from the desired number of 360 CHs per round. On the one hand, in Figure 4.b, one may underline that almost 60% of the RLP rounds have less than 100 available CHs, what explains the CH shortage of RLP. In Figure 4.c, on the other hand, shows that more than 90% of the rounds of HRLP have more than 300 CHs.

Notice that the CH shortage of RLP impacts on its capability of reacting to a connectivity decrease, remarkably by its inability to reintroduce new CH nodes. This can be observed in Figure 3.a when attaining for the small gaps between data series. In contrast, the CH recovery employed on HRLP is capable of sustaining CH nodes throughout the network operation.




(a) *Lifetime* for RLP. (b) *Lifetime* for LEACH. (c) *Lifetime* for HRLP.

Fig. 3. The *lifetime* results. Each data series corresponds to a distinct recovery threshold $\gamma$.



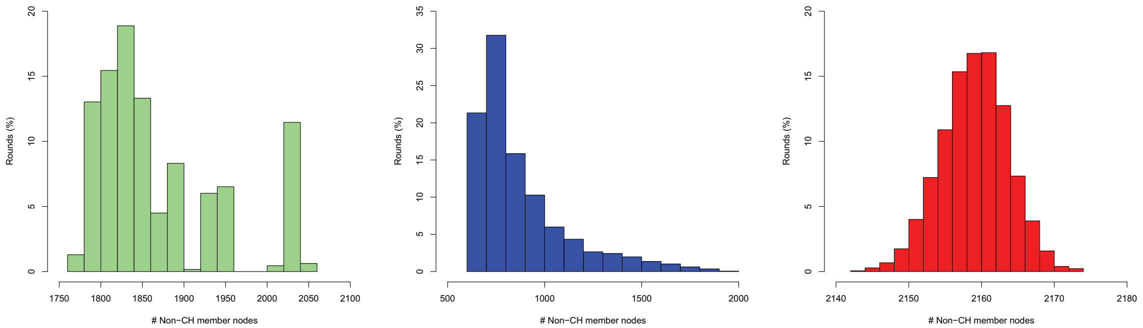

(a) CH histogram of LEACH. (b) CH histogram of RLP. (c) CH histogram of HRLP.

Fig. 4. The histograms related to the CH availability in the simulated scenario.

To validate the benefits of the topology-awareness provided by the community detection-based approach employed in our proposal, Figure 5 displays the results with respect to the quantity of non-CH nodes connected to clusters in the simulated scenarios for the evaluated protocols. Each bar of the histogram represents the percentage of rounds with the related range of non-CH members. It is worthy mentioning that the average number of nodes connected to the CHs in every round of RLP, LEACH and HRLP were, respectively, 876.34, 1869.41 and 2159.06. Additionally, the average number of nodes per CH in each round were, respectively, 8.18, 4.83 and 5.88. Therefore, the proposed HRLP presented better CH distribution than RLP and LEACH, which consequently may lead to a better balance in the network load.

## 7. Final Remarks

In this paper, we introduced a hybrid CH election algorithm designed to community detection-based routing protocols in WSNs. The proposed strategy is an extension of the Routing Label Propagation (RLP), here named Hybrid Routing Label Propagation (HRLP). The HRLP, as its precursor, is a community detection-based protocol that relies on a computationally cheap and distributed probabilistic-based CH recovery procedure that improves the network connectivity by electing CHs on demand inside the communities. Additionally, we provided a performance evaluation of our proposal and a comparison to other protocols by considering a large-scale WSN scenario. The results evinced the improvements achieved by HRLP by means of the network connectivity and lifetime metrics. As future research




(a) Non-CH histogram of LEACH. (b) Non-CH member histogram of RLP. (c) Non-CH member histogram of HRLP.

Fig. 5. The histograms related to the Non-CH distribution in the simulated scenario.

we intend to evaluate the performance of the HRLP in a greater variety of large-scale scenarios regarding the latency and data delivery reliability metrics.

## Acknowledgements

The authors would like to thank to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP Proc.: 2015/21660-4, 2015/18580-9) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CNPq id: 448614/2014 6) for their financial support.

## References

1. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.. Wireless sensor networks: a survey. *Computer Networks* 2002;**38**(4):393 – 422.
2. Santi, P.. Topology control in wireless ad hoc and sensor networks. *ACM Comput Surv* 2005;**37**(2):164–194.
3. Pantazis, N., Nikolidakis, S., Vergados, D.. Energy-efficient routing protocols in wireless sensor networks: A survey. *Communications Surveys Tutorials, IEEE* 2013;**15**(2):551–591.
4. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.. Energy-efficient communication protocol for wireless microsensor networks. In: *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*. 2000, p. 10 pp. vol.2–.
5. Mamalis, B., Gavalas, D., Konstantopoulos, C., Pantziou, G.. Clustering in wireless sensor networks. *RFID and Sensor Networks: Architectures, Protocols, Security and Integrations, Y Zhang, LT Yang, J Chen, eds* 2009;:324–353.
6. De Paulo, M., Nascimento, M., Rosset, V.. Rlp: A community detection-based routing protocol for wireless sensor networks. In: *Network Computing and Applications (NCA), 2014 IEEE 13th International Symposium on*. 2014, p. 237–244.
7. Handy, M., Haase, M., Timmermann, D.. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In: *Mobile and Wireless Communications Network, 2002. 4th International Workshop on*. 2002, p. 368–372.
8. Liu, Y., Luo, Z., Xu, K., Chen, L.. A reliable clustering algorithm base on leach protocol in wireless mobile sensor networks. In: *Mechanical and Electrical Technology (ICMET), 2010 2nd International Conference on*. 2010, p. 692–696.
9. Chan, H., Perrig, A.. Ace: An emergent algorithm for highly uniform cluster formation. In: *Wireless Sensor Networks: First EuropeanWorkshop, EWSN 2004, Berlin, Germany, January 19-21, 2004. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004, p. 154–171.
10. Wen, C.Y., Sethares, W.A.. Automatic decentralized clustering for wireless sensor networks. *EURASIP J Wirel Commun Netw* 2005;**2005**(5):686–697.
11. Brust, M.R., Andronache, A., Rothkugel, S., Benenson, Z.. Topology-based clusterhead candidate selection in wireless ad-hoc and sensor networks. In: *Communication Systems Software and Middleware, 2007. COMSWARE 2007. 2nd International Conference on*. 2007, p. 1–8.
12. Watts, D.J., Strogatz, S.H.. Collective dynamics of small-worldnetworks. *nature* 1998;**393**(6684):440–442.
13. Ngo, H., Lee, Y.K., Lee, S.. Mepa: A new protocol for energy-efficient, distributed clustering in wireless sensor networks. In: *Wireless Communication Systems, 2007. ISWCS 2007. 4th International Symposium on*. 2007, p. 40–44.
14. Raghavan, U.N., Albert, R., Kumara, S.. Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E* 2007;**76**:036106.
15. Boulis, A., et al. Castalia: A simulator for wireless sensor networks and body area networks. *NICTA: National ICT Australia* 2011;.